

COVID-19 Case Surveillance Public Use Data with Geography Utility Summary

Users should consider the level of completeness, including suppression levels when planning their analyses and use of public datasets. Privacy protections will suppress field values to reduce reidentification risks. Completeness varies by jurisdiction (i.e., state, local, and territorial) and time period. Variables are consistently coded to the value “Unknown” when jurisdictions specify in the case data submitted to CDC that the value is unknown, the value “Missing” when jurisdictions do not provide a value, and the value “NA” when the value is suppressed as part of privacy protections.

Dataset version: 2021-12-20

Total records in dataset: 40,911,354

Quick Summary

	all_fields	quasi_fields
total_fields	19	8
total_records	40,911,354	40,911,354
total_cells	777,315,726	327,290,832
missing_fields	169,061,265	24,563,369
missing_pct	22%	8%
complete_fields	608,254,461	302,727,463
complete_pct	78%	92%
unknown_fields	33,822,915	18,570,538
unknown_pct	4%	6%
suppressed_fields	23,256,460	20,403,404
suppressed_pct	3%	6%
available_fields	551,175,086	263,753,521
available_pct	71%	81%

Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
case_month	12	0.0%	0	0.0%
res_state	1,020	0.0%	0	0.0%
res_county	2,852,036	7.0%	0	0.0%
age_group	458,157	1.1%	363,196	0.9%
sex	1,263,631	3.1%	47,215	0.1%
race	6,348,079	15.5%	2,836,301	6.9%
ethnicity	7,550,933	18.5%	2,415,089	5.9%
death_yn	1,929,536	4.7%	18,901,568	46.2%
records_with_any_field	11,046,265	27.0%	19,979,428	48.8%